

El Proceso de Modelado de la Ingeniería del Conocimiento como Método de Optimización en la Resolución de Problemas

Claudia Jiménez Ramírez y Hernán Darío Álvarez Zapata

UNIVERSIDAD NACIONAL DE COLOMBIA, Medellín.
{csjimene ; hdalvare} @unal.edu.co

Recibido para revisión May-2006, aceptado Jun-2006, versión final recibida Jun-2006

Resumen: Se propone el uso de modelos conceptuales de la Ingeniería del Conocimiento para la comprensión y análisis de problemas de carácter científico y su dominio, sin importar que el modelo de la solución no sea un modelo computacional. Se ilustrará el uso de estos modelos mediante un ejemplo.

1 INTRODUCCIÓN

Una de las mayores dificultades para la resolución de problemas tiene que ver con la comprensión del mismo y de su entorno para poder plantear la mejor solución. En el presente trabajo se propone, como método de optimización, que la resolución de problemas de carácter científico se considere como un proceso de elaboración de modelos de la Ingeniería del Conocimiento que faciliten la comprensión del dominio del problema y la comunicación entre los analistas expertos, o con los interesados, antes de llegar a un modelo de la solución, expresado en un lenguaje matemático algebraico.

Metodologías recientes de la Ingeniería del Conocimiento, proponen la elaboración de varios modelos para representar todos los aspectos que deben considerarse sobre el dominio del problema y que conducen a la mejor solución informática, de modo parecido a las metodologías modernas de la Ingeniería del Software. En este trabajo se presenta cómo los modelos de la Ingeniería del Conocimiento pueden usarse en otras áreas o disciplinas que no tengan como fin último, o modelo de la solución, un modelo de computador. Se mostrará que el enfoque sistémico, con su forma de representación y sus lenguajes semiformales, puede ser muy útil para la comprensión de un problema y la representación de su entorno, cuando la notación algebraica convencional no es procedente o conveniente para representar todas las propiedades relevantes o las restricciones de un dominio particular.

2 PLANTEAMIENTO DEL PROBLEMA

Una investigación científica inicia con el planeamiento, en lenguaje natural, del problema por resolver y luego de

un análisis cuidadoso del entorno, se plantea una solución que se suele especificar formalmente mediante un lenguaje algebraico. Este último lenguaje, por ser universal y más formal (pues no da pie a tantas malas interpretaciones), tiene la propiedad de generar modelos que pueden ser validados por los expertos en los asuntos en cuestión y compartido con la comunidad científica. Sin embargo, no todos los aspectos relevantes del dominio del problema pueden representarse por medio de fórmulas algebraicas, como cuando se requiere especificar un proceso que involucre no sólo cálculos matemáticos, sino ciertas características estructurales en el entorno para llegar a un planteamiento completo de la solución. En estos casos, la traducción directa del modelo verbal, al modelo de la solución en un lenguaje algebraico implicaría una pérdida de información que dificultaría la verificación de que el modelo de diseño de la solución si es el apropiado para la resolución de un problema dado.

Por otro lado, las especificaciones en lenguaje natural, por lo regular, son ambiguas y vagas permitiendo diferentes interpretaciones de lo que se quiere decir en un texto. Por esto, antes de comenzar a diseñar la solución es necesario verificar si hubo una buena interpretación del problema mediante el uso de modelos gráficos que sinteticen y aclaren detalles del mismo; evitando malos entendidos que puedan conducir a la no satisfacción de las necesidades planteadas. La eliminación de los errores, lo más rápido posible, es la clave para aumentar la productividad y optimizar el proceso de desarrollo de una investigación.

3 ESTADO DEL ARTE

En Inteligencia Artificial, se considera la Ingeniería del Conocimiento, como la disciplina que orienta en el

planeamiento y la programación de las tareas para la construcción de sistemas basados en conocimiento (KBS, por sus siglas en inglés). Actualmente, existe el consenso de considerar el desarrollo de KBS como el desarrollo de modelos de computador con capacidades para la resolución de problemas comparables a las realizadas por expertos calificados en el dominio [Studer, Benjamins y Dieter (1998)].

Como en la Ingeniería de Software, también se han desarrollado varias metodologías y lenguajes de especificación de KBS con el intento de convertir el desarrollo de KBS en una disciplina ingenieril. Entre ellas, se pueden mencionar a CommonKADS y a PROTEGE-II que explota la noción de ontologías [Schreiber (2000)].

La metodología CommonKADS considera una colección de modelos para el análisis y diseño de la solución a un problema de representación o de descubrimiento de conocimiento, del mismo modo que se sugiere en las metodologías de la Ingeniería de Software. En ella, se distinguen 6 modelos y cada uno permite representar aspectos específicos del KBS: el modelo del dominio o de la organización, el modelo de los agentes (humanos, dispositivos o software) que intervienen en la solución del problema, el modelo de las tareas requeridas, el modelo de conocimiento (conformado por los datos y las reglas o restricciones aplicables en ese contexto), el modelo de comunicación, que muestra cómo se comunican los agentes para llevar a cabo las tareas encomendadas y, como un modelo de más bajo nivel de abstracción, el modelo de diseño de la solución que incorpora detalles no funcionales. Es decir, aquellos detalles que tienen que ver con la eficiencia y la adopción de ciertos recursos tecnológicos necesarios para la solución del problema.

Las metodologías modernas de la Ingeniería de Conocimiento, comparadas con las metodologías de la Ingeniería del Software, aportan el concepto de agente y el modelo de conocimiento como una ampliación del modelo de datos u objetos del dominio para incluir reglas y procedimientos (operaciones) aplicables en él. Por esto, permiten modelar mayores aspectos relevantes durante la conceptualización del problema. El lenguaje de representación gráfico puede ser el mismo usado en la Ingeniería del Software, buscando estandarización y mayor universalidad. Actualmente, en esta disciplina, se ha generalizado el uso del lenguaje UML (*Unified Modeling Language*).

El lenguaje unificado UML fue propuesto entre fines de los ochenta y principios de los noventa, como una notación principalmente gráfica de la que se puede valer cualquier método para expresar los modelos conceptuales y de diseño de sistemas informáticos y considerado el estándar por la OMG [Grady, Rumbaugh y Jacobson (1998)]. Actualmente, va en su segunda versión.

Otros modelos conceptuales alternativos que son utilizados en la Ingeniería del Conocimiento también

pueden ser utilizados para la especificación de un problema científico. Actualmente, en muchos establecimientos educativos de distintos sitios del mundo se está enseñando a utilizar herramientas para representar el conocimiento como los mapas conceptuales o la "V epistemológica" para que el estudiante modele lo visto en un capítulo o tema de estudio. Con estos modelos conceptuales, el profesor, a su vez, puede evaluar si el estudiante ha comprendido los nuevos conceptos impartidos [Moreira (1997)].

Los mapas conceptuales permiten a los individuos entender fenómenos y hacer inferencias. Se trata de modelos de trabajo que incluyen representaciones proposicionales e imágenes, partes expresables verbalmente y perceptibles del modelo mental, respectivamente para la comunicación comprensiva entre individuos [Johnson-Laird (2000)].

4 PLANTEAMIENTO DE LA SOLUCIÓN

En las fases iniciales del análisis del problema, los científicos construyen modelos mentales expertos acerca del fenómeno en cuestión. Los modelos mentales expertos "no se ven", pues están en las mentes de los científicos y por eso, deben definirse explícitamente para ser discutidos con pares o con los interesados en la resolución del problema. No es recomendable pasar directamente del planteamiento del problema a especificaciones formales, en un lenguaje algebraico o a un lenguaje de programación, pues no sólo puede dificultar la comprensión de la solución propuesta, sino que pueden quedar por fuera algunos aspectos estructurales o restricciones en el dominio o universo del discurso.

Los modelos propuestos en CommonKADS pretenden cubrir aspectos funcionales, estructurales y dinámicos del fenómeno bajo estudio y de esta manera no omitir aspectos relevantes que incidan en la calidad de la solución. Dichos modelos constituyen vistas parciales del mismo dominio que ayudan a reducir la complejidad, separando un pequeño número de cosas importantes para tratarlas a la vez. Sin embargo, son vistas no disjuntas pues para tratar un aspecto, muchas veces, es necesario especificar elementos ya descritos o especificados en otros modelos; aunque no necesariamente con el mismo detalle o bajo la misma óptica. De modo que se deberá chequear que no existan contradicciones o inconsistencias entre los modelos creados para la representación de todos los aspectos relevantes en la resolución de un problema.

A continuación, se describen brevemente cada uno de los modelos conceptuales del CommonKads y se ilustra su utilización, con un ejemplo hipotético: una fábrica de comida precocida congelada que está altamente preocupada por el bajo nivel de ventas de algunos de sus productos. No sabe si son los precios o el sabor de los mismos. Por eso, el departamento de Investigación de

Mercados debe dar luces a las directivas sobre lo que está pasando para que se tomen las medidas correctivas.

4.1 Modelo del Dominio

Describe todas las funciones llevadas a cabo por cada unidad organizacional o por los agentes en un dominio determinado. Se incluyen las deficiencias del sistema actual y las oportunidades identificadas para mejorar los procesos con el KBS. Esto último se puede hacer con la construcción de una matriz DOFA y por esto no se muestra gráficamente. También el organigrama es parte de este modelo.

La especificación gráfica del modelo funcional o de comportamiento, en el dominio en cuestión, puede llevarse a cabo con el Diagrama de Casos de Uso, considerando cada uno de ellos como un proceso o actividad de un agente para lograr los objetivos que le son enunciados.

Los conceptos principales que permite representar el Diagrama de Casos de Uso en UML 2.0 son los casos de uso, los actores y las relaciones entre ellos [Novak (1988)]. Un caso de uso describe una secuencia de acciones que proveen a los actores un valor agregado o medible y son dibujados como elipses. Los actores son personas, unidades organizacionales o sistemas externos que juegan un papel relevante en una o más de las acciones del sistema modelado y se dibujan con un ícono de una figura humana. El concepto de actor se puede asimilar al concepto de agente de la Ingeniería del conocimiento y por eso, para diferenciar los agentes humanos de los no humanos, se usa una caja con el nombre del agente y con el ícono en el interior. Las asociaciones entre casos de uso y actores son indicadas por las líneas. En la Figura 1, se muestra el Diagrama de Casos de Uso para los procesos o funciones, de alto nivel de abstracción, para la dependencia de Investigación de Mercados.

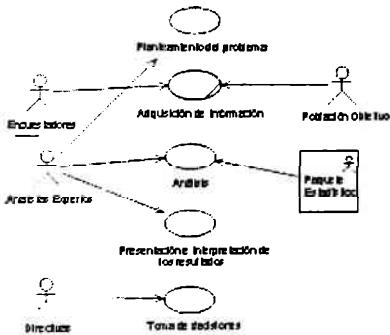


Figura 1: Diagrama de Casos de Uso en la resolución de problemas de investigación de mercados

Actualmente existen gran variedad de herramientas computacionales, comerciales y de distribución libre que permiten la creación de diagramas, usando lenguaje UML y que permiten completar los modelos conceptuales con texto para especificar detalles de cada elemento. En la Figura 2, se muestra la forma para entrar los detalles del caso de uso "Adquisición de Información".

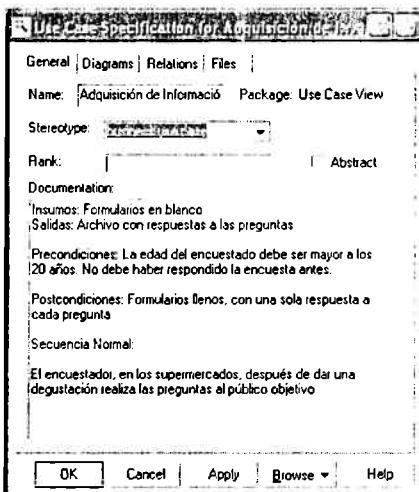


Figura 2: Forma para la documentación de la adquisición de información

4.2 El Modelo de Tareas

El Diagrama de Actividades permite representar, además de los procesos de la organización, los flujos de datos y materiales necesarios como entradas o salidas de un paso, o proceso, que se suelen representar con cajas cuadruplicadas o con bordes redondeados. También permite describir el evento que da inicio al proceso y los eventos terminales del proceso. En este diagrama, la actividad representa un paso de un proceso o un proceso completo, admitiendo la descomposición de una actividad compleja, en otras más simples y se dibuja como un círculo o caja redondeada. Además, permite modelar la lógica compleja de un sistema o proceso pues se consideran puntos de decisión o bifurcación y se admiten actividades en paralelo. También se admite definir carriles para mostrar como se distribuyen las actividades o tareas entre los agentes o actores. En la Figura 3, se ilustra el diagrama de actividades generales (admiten descomposición) para el ejemplo.

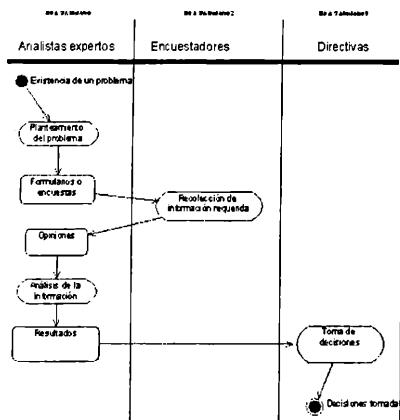


Figura 3: Diagrama de Actividades de Investigación de Mercados

4.3 El Modelo de Agentes

Este modelo describe las características de los agentes que, como se dijo antes, pueden ser humanos, dispositivos o programas de computador. De cada agente se especifica, su tipo, sus responsabilidades y tareas, su grado de autonomía y las restricciones que pueda tener para actuar en el entorno. Para la especificación de estas características, se usa la plantilla de trabajo AM-1 en la metodología CommonKads [Schreiber (2000)] que se muestra en la Tabla 1 y usada para mostrar las características del agente "analista" en el problema de investigación de mercados.

Tabla 1: Hoja de Trabajo AM-1. Agentes

Nombre	Analista
Tipo	Humano
Ubicación	Oficina Investigación de Mercados
Involucrado en	Planteamiento del problema, análisis
Se comunica con	Encuestadores, Directivas
Conocimientos	Mercado, Ventas, Estadística
Otras competencias	Creatividad, recursividad
Restricciones	No tener nexos familiares con encuestadores

4.4 El Modelo de Comunicaciones

En éste se especifican las interacciones entre los diferentes agentes para llevar a cabo una tarea conjunta. Específica, por ejemplo, el tipo de información que es

intercambiado y cuál agente inicia la interacción. También se pueden especificar los tiempos de respuesta en cada interacción. Un ejemplo de este modelo, que se realizó con el Diagrama de Secuencias de UML, aparece en la Figura 4. Allí se especifica, por ejemplo, que se espera que transcurran dos semanas para la entrega de los formularios diligenciados por parte de los encuestadores.

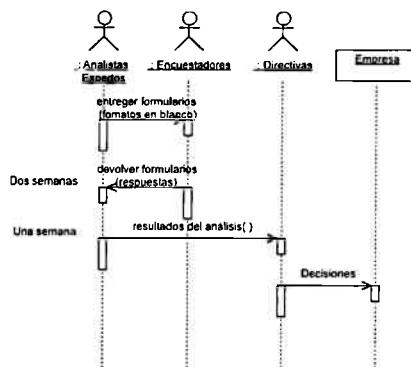


Figura 4: Comunicación entre Agentes

4.5 El Modelo del Conocimiento

Permite definir el conocimiento necesario para lograr los objetivos especificados en el Modelo de la Organización y de las Tareas. Es el corazón de la metodología CommonKads pues permite representar el conocimiento para la resolución de problemas empleado por un agente en una tarea. Este modelo estructura, en un único esquema, las tres capas de conocimiento requeridas para resolver una tarea particular: una *capa del dominio* que incluye los conceptos del dominio y la teoría requerida para resolver un problema, la *capa de inferencia* que describe cómo razonar y usar el conocimiento del dominio y, la tercera capa, la *capa de la tarea*, representa una estrategia particular para lograr los objetivos que se persiguen con una tarea determinada.

En la Figura 5, se muestra el modelo de experticia para la tarea "prueba de hipótesis para la diferencia de proporciones" que permite evaluar si existen diferencias significativas en la proporción de respuestas positivas en las características evaluadas de los productos, de acuerdo con el estrato o con el sexo de las personas interrogadas.

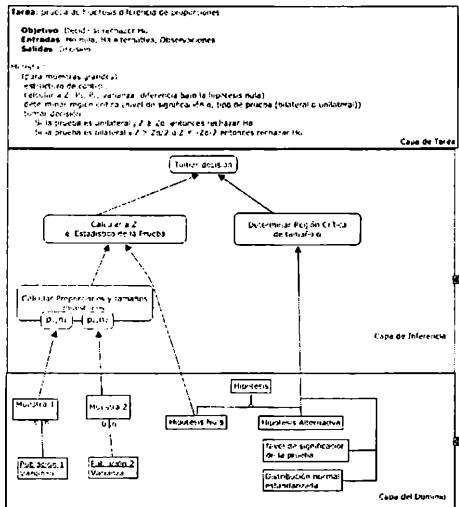


Figura 5: Modelo de Experticia para pruebas de hipótesis de proporciones

4.6 Los mapas conceptuales

Los mapas conceptuales permiten pensar sobre las operaciones mentales que se dan en el cerebro ya que mediante la elaboración y reclaboración los mismos, el individuo o el grupo ejercitan sus procesos metacognitivos al reflexionar sobre los resultados del propio aprendizaje y comprender el procedimiento implícito que se utiliza para obtener ese conocimiento. Permiten tomar conciencia de varias cosas útiles a la hora de avanzar en una investigación: a) ¿Cuánto sé? b) ¿Cómo tengo estructurado lo que sé? y c) ¿Qué me falta saber en relación con un determinado cuerpo de conocimiento?. Obviamente, la respuesta estas tres preguntas permiten armar una estrategia de lecturas-discusiones que faciliten la definición-delimitación del área problemática y dentro de ella, el problema a tratar en la investigación. El mismo conjunto de preguntas puede utilizarse durante el estudio de las herramientas que se requieren usar para la validación de las hipótesis de investigación. Al igual que con las Taxonomías, los Mapas Conceptuales permiten una "Hoja de Ruta" sobre la cual se pueden trazar caminos de solución (Hipótesis) al Problema e incluso deducir Objetivos a desarrollar para lograr la validación de la Hipótesis.

De acuerdo con la definición original de Novak (1988) el Mapa Conceptual contiene los siguientes elementos:

- Conceptos.** Se entiende por concepto el conjunto de atributos que una persona asocia con el símbolo que representa a una clase (o agrupación) de objetos, eventos o ideas.
- Estructura.** Se constituye por las relaciones significativas entre dos o más conceptos. Se forma mediante palabras de enlace y constituyen una unidad semántica cuyo significado es contextual.
- Palabra-enlace.** Son las palabras que unen los conceptos y señalan el tipo de relación existente entre ellos.
- Jerarquías.** En los Mapas Conceptuales los conceptos se disponen en orden de importancia o de inclusión formando jerarquías o niveles de subordinación, en los que los conceptos más generales ocupan los lugares superiores de la estructura gráfica, hasta llegar a la base en la que aparecen los conceptos más específicos y si es del caso los ejemplos.
- Categorías.** Son agrupaciones de conceptos que poseen características o elementos comunes y se denominan con un término conceptual general.
- Enlaces cruzados.** Son las relaciones que establece entre dos o más conceptos de categorías diversas.
- Ejemplos.** Son los eventos u objetos reales ó simbólicos que representan el término conceptual. Los ejemplos ponen de manifiesto la comprensión de los nuevos conceptos ya que su elaboración requiere un proceso mental para aplicar el conocimiento en casos concretos.

En la Figura 6, se muestra el mapa conceptual de una prueba de hipótesis para la diferencia de proporciones.

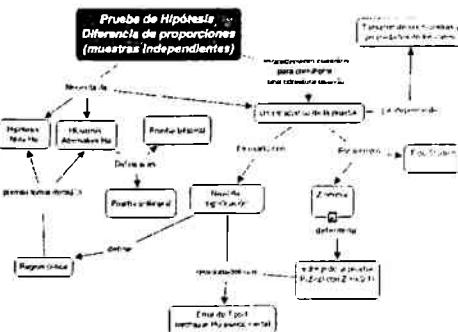


Figura 6: Mapa Conceptual para pruebas de hipótesis de proporciones

Un mapa conceptual es muy parecido al Diagrama de Clases de UML pues éste permite representar las distintas clases de objetos (conceptos) y sus relaciones. Tiene mayor poder expresivo puesto que permite especificar todas las propiedades (atributos y operaciones) aplicables a los conceptos, dentro de la caja que representa una clase, y distingue visualmente los tres tipos de relaciones entre las clases: la composición o relación "parte de" que se adorna con un rombo junto a la clase agregada, la relación de especialización o generalización, que se adorna con un triángulo, y la asociación que equivale a un enlace cruzado del mapa conceptual. También permite definir la cardinalidad mínima y máxima de dichas relaciones. Pero no permite representar las instancias o ejemplos por no considerarse conceptos. El modelo del dominio, como capa del modelo de experticia, en la Figura 5, fue construido usando el diagrama de clases de UML Allí se presenta una relación de asociación entre las muestras y las poblaciones con una cardinalidad $0 \dots n$ y se muestra una relación de generalización, o especialización, entre la hipótesis y las subclases hipótesis nula y alternativa.

5 CONCLUSIONES

Los modelos propuestos en la Ingeniería del Conocimiento pueden ser muy útiles en la comprensión y análisis de cualquier tipo de problema como se corrobora con el ejemplo presentado y no tienen por qué ser exclusivos de los informáticos.

Los modelos conceptuales son aplicables cuando el grupo de investigación intenta dirimir sobre aspec-

tos relacionados con el dominio del problema o incluso cuando está en la etapa de proponer expansiones a conceptos existentes o un concepto totalmente nuevo.

Los modelos conceptuales para la resolución de un problema particular facilitan la búsqueda de la solución óptima pero su utilidad dependerá no sólo de las bondades o propiedades de los modelos empleados, sino de las habilidades para modelar de los expertos. Por eso, se requiere del dominio del lenguaje de modelado por parte de los investigadores, tanto en sus aspectos semánticos como sintácticos.

REFERENCIAS

- Grady, B., Rumbaugh, J. y Jacobson, I. (1998). *The Unified Modeling Language User Guide*, ISBN 0-201-57168-4, Addison-Wesley.
- Johnson-Laird, P. (2000), *Mental Models in Reasoning*, García-Madruga et al., UNED, Madrid, chapter The current state of mental model theory.
- Moreira, M. (1997), 'Aprendizaje significativo: fundamentación teórica y estrategias facilitadoras'. Monografías del Grupo de Enseñanza de la Universidad Federal do Rio Grande do Sul. Brasil.
- Novak, J. y Gowin, D. (1988). *Aprendiendo a Aprender*, Ediciones Martínez Roca S.A.
- Schreiber, M. (2000). *Knowledge Engineering and Management: The CommonKADS Methodology*. MIT Press.
- Studer, R., Benjamins, R. y Dieter, F. (1998). 'Knowledge engineering: Principles and methods', *Data & Knowledge Engineering* pp. 161-197.